The fundamental problem in psycholinguistics is simple to formulate: what happens when we understand sentences? Most of the other problems would be half-way solved if only we had the answer to this question; and potential answers to it, of course, have accumulated rapidly ever since George Miller (125) introduced psychologists to the ideas of transformational grammar. The results of the initial phases of experimentation are well known (see, for example, 73, 83, 160); and since Fillenbaum (52) has recently contributed a wide-ranging survey of the area, including language acquisition and its biological bases, it is appropriate in the present review to concentrate on comprehension and its cognate problems.

Because meaning depends on words and syntax, there is a general consensus that comprehension involves processing both of them and combining the meanings of words according to the perceived syntactic relations between them. But comprehension is only a start: once an individual understands a sentence he may react to it. He may try to verify a statement, or to answer a question, or to carry out a command. He may simply remember the sentence or one of its implications. Many of his tangible responses are of immediate intrinsic interest to a psycholinguist. But the responses do have a further use. If an individual can decide that a sentence is true, then he is likely to have come to an understanding of it, and his performance may therefore be very revealing about this process. Such indirect approaches to the study of comprehension are extremely useful because there is no simple veridical index of when a sentence has been understood.

The literature continues to grow faster than knowledge, and in trying to epitomize its findings and foibles, we shall follow the progress of a sentence as its syntax is parsed, its word meanings retrieved from lexical memory, and its sense ultimately understood. We shall examine such matters as how sentences are verified, and how questions are answered. Finally, we shall turn to the ways in which sentences and larger units of discourse are remembered.

EXPERIMENTAL PSYCHOLINGUISTICS

P. N. Johnson-Laird

University of Sussex, Brighton, England
THE PERCEPTION OF SYNTAX

It was implicit in the initial phase of experimental psycholinguistics that the process of perception paralleled the levels of analysis in transformational grammar. Yet it was clear that there would have to be considerable interplay between these levels because purely semantic factors were known to affect the intelligibility of speech. Nevertheless, the pioneers believed that the linguistic distinction between surface and deep structure was likely to be reflected in the organization of the perceptual system, with one set of procedures designed to parse a sentence into its immediate constituents—nounphrases, verbphrases, etc.—and another set of procedures retrieving the underlying structure of a sentence, including such relations as its subject and object. The distinction was there from the start and it has proved vital to most of the work in the area, particularly the sustained investigations into perception carried out by Bever, Fodor, Garrett, and their colleagues at MIT.

The MIT group has made considerable use of a technique in which listeners are required to locate a click presented to one ear with respect to a sentence presented to the other ear. The subjects usually have to write down the sentence and then to indicate where the click occurred within it; however, comparable results appear to be obtained when they have merely to judge whether a click occurred in the same place in two sentences (cf. the unpublished MIT study cited in 84). The first studies with the technique, reviewed by Bever (14), indicated that the click was often perceptually relocated so as to minimize the number of surface constituents that it interrupted. It was as though extraneous noises were not permitted to violate the integrity of major constituents and were consequently preposed, or postposed, into the boundaries between them. The results of more recent studies, however, suggest that whatever it is that determines click location, it is not surface structure.

The Reality of Deep Structure

Reber & Anderson (148) certainly found a regression of clicks towards the major surface boundary within sentences. But they had purposely confounded this boundary with the midpoint of the sentences, and a similar regression occurred with both random strings of words and strings made up of bursts of white noise. Indeed, a distinct response bias towards this position was evident in the performance of a control group instructed to respond to "subliminal" but in fact nonexistent clicks. There is evidently a similar bias towards locating clicks within words rather than between them (109) which also tells against the hypothesis since words are surface constituents. Such findings naturally suggest that nonlinguistic attentional factors may be primarily responsible for the perceptual location of clicks. But although the MIT group subsequently failed to detect effects of surface structure, they have regularly reported another phenomenon that seems inexplicable without resort to linguistic principles. Clicks are consistently relocated so that they fall into or towards a boundary between the underlying clauses of sentences—even, Bever (14) claims, when such boundaries are unmarked in surface structure. A particularly striking finding involved a contrast between two superficially similar sorts of complement (18). With a nounphrase complement, an entire underlying clause is treated
like a nounphrase, e.g. "The corrupt police can’t bear criminals to confess very quickly." A click actually located in the word "criminals" would often be reported to have occurred between "bear" and "criminals," i.e. at the point corresponding to the boundary between the two underlying clauses. With a verbphrase complement, on the other hand, there is no longer a distinct boundary between clauses, e.g. "The corrupt police can’t force criminals to confess very quickly." Here "criminals" serves as the object of "force" as well as the subject of "confess"; and a click located in the word "criminals" no longer showed any systematic tendency to be preposed. It is debatable whether or not there are concomitant differences in the surface structure of these two sorts of sentence. Chapin and his colleagues (30) were unhappy about this aspect of the materials and undertook a more direct test of Bever’s conjecture. They found that clicks placed midway between a major surface boundary and a subsequent underlying clause boundary were attracted to the surface boundary; but when the clause boundary preceded the major surface boundary, locational errors were more or less equally divided between them. It seems likely that the technique will have to be refined in order to resolve these inconsistencies. One feasible modification would be to utilize the paradigms of signal detection, setting listeners to detect the presence of a faint click in a specified region of a sentence or forcing them to choose between a number of possible alternative locations. These procedures may show that there are genuine fluctuations in perceptual sensitivity during the syntactic analysis of sentences. Such a view is plausible because sensitivity to the location of a switch in speech from one ear to the other is greatest at boundaries between clauses (183), and similarly reaction times to clicks are faster when they occur in clause boundaries (84). Meanwhile, however, what are we to infer from the rather muddled history of clickology?

The story told by Bever (16) has considerable attraction. He argues that surface structure has a less direct perceptual reality than underlying structure. Surface structure is a derived construct, and it takes time to mature in the mind of the perceiver. Its effects on perceptual tasks such as the location of clicks occur only when there is sufficient delay for its mental derivation to take place. Ordinarily, when an individual listens to a sentence, he is gathering information about its underlying structure, and it is at the end of each clause that such structure is assigned—hence the relocation of clicks to clause boundaries. Bever goes on to argue that clauses are more directly perceived than individual words, since listeners recognize a target word in a list of sentences more quickly when they are told the sentence in which it will occur, even though it is the first word of the sentence (15). Moreover, Savin & Bever (154) argue that words are more directly perceived than phonemes. They found that listeners recognize a target phoneme in a list of syllables more quickly when they are told the syllable in which it will occur, even though it is the first phoneme of the syllable. This topsy-turvy conception, from which it seems to follow that deep structures are more directly apprehended than phonemes, makes it difficult to explain such everyday phenomena as the detection of mispronunciations, neologisms, and solecisms. Yet is is a useful supplement to the more conventional order of processing—from sound through syntax to sense. It embodies the important truth that the perceiver does not come empty-handed to his task. He is
likely to have a number of conjectures about the meaning of a sentence, derived from his immediate or general semantic knowledge. He is likely to possess a variety of perceptual heuristics to aid him in the assignment of underlying structure, e.g. the occurrence of a noun-verb-noun sequence normally signifies the underlying roles of actor-action-object (14).

If a conventional order of processing is permitted to augment Bever's scheme, then it is feasible that a perceiver also exploits cues in the surface structure of a sentence in order to recover its underlying structure (55). Such cues appear to divide into two categories: the presence of grammatical formatives indicating explicit structure, and the presence of certain lexical items suggesting definite hypotheses about structure. Particular exemplars of both sorts of cue have been investigated using the phoneme-monitoring task devised by Foss (59). The technique rests on the assumption that all aspects of perceptual analysis are carried out to some extent by the same central processor; hence the time taken to recognize a target phoneme provides a useful index of the load placed on the processor by the syntax of a sentence. The general assumption appeared to be justified by the finding that responses were slower when a target followed a relatively infrequent word, although the effect only occurs with adjectives (24). One salient cue to underlying structure is the presence of relative pronouns in embedded sentences. For example, "The car that the man whom the dog bit drove crashed" seems easier to understand than "The car the man the dog bit drove crashed." Despite an early setback (62), Foss and Hakes and their co-workers have established that responses to a target phoneme in an embedded sentence are faster when it contains relative pronouns (77, 78). Hakes (76) has demonstrated in the same fashion that signaling the occurrence of a complement by the presence of "that" makes a sentence easier to parse.

The effects of lexical cues to underlying structure have been examined only with verbs, perhaps because they differ so markedly in the number and variety of structures they may take as objects. Verbs that can take complements should cause more difficulties than verbs that can take only simple direct objects, since whenever a complement verb occurs a greater number of structural hypotheses have to be tested (56). However, Hakes (75) failed to detect any effect of verb complexity upon a phoneme-monitoring task. A possible explanation for this failure is suggested by the results of an experiment carried out by Holmes & Forster (85). They used a technique invented by Forster (57) in which the words of a sentence are superimposed one after the other in a rapid serial visual presentation. Under this regime, the perception of sentences was influenced by the sort of verb they contained. Sentences with simple transitive verbs were easier to perceive accurately than sentences with nounphrase complement verbs, and when such a verb was followed by an actual complement, e.g. "Bill wanted his father to leave on the last bus," the sentence was still harder to perceive. A very different pattern of results emerged with verbphrase complement verbs. The presence of such a verb, e.g. "The judge warned the youth about the dangers of drugs," made a sentence more difficult to perceive, but this difficulty vanished completely when such a verb was followed by an actual complement, e.g. "His mother allowed the boy to swim in the pool." This surprising facilitation is probably due to the fact that verbphrase complements are much more
constrained than nounphrase complements: they must contain an explicit subject, which is almost invariably animate. Since Hakes (75) had used an identical sentence frame for presenting both the simple and the complement verbs, his technique may have been insensitive to such subtle differences between them. This explanation may also account for a recent failure to find an effect of verb complexity on judgments of grammaticality (71).

The Unreality of Deep Structure

For obvious historical reasons most psycholinguists have maintained a continued intellectual loyalty to the basic tenets of transformational grammar. Yet if Chomsky and his colleagues had not postulated the existence of deep structure, psycholinguists might never have invented it. The studies of click location show that clicks migrate sometimes to the boundaries between underlying clauses and sometimes to boundaries between major surface constituents. A simple explanation for this apparent inconsistency is that the technique may be primarily sensitive, not to the assignment of syntax, but to the amalgamation of meaning. The bulk of semantic processing obviously takes place at the end of clauses and at the end of complicated nounphrases. Just such nounphrases attracted clicks in the study carried out by Chapin and his colleagues (30). Similarly, the effects of different sorts of verb on rapid serial visual perception may be attributable to semantic factors: verbphrase complements almost invariably have animate agents, but there is no comparable prediction for nounphrase complements.

This tentative conjecture is not a covert plea for a particular conception of linguistics, but is an argument for abandoning the idea that the deep structure of a sentence is separately perceived. It suggests instead that there is a single integrated analysis of syntax aimed at determining the meaning of a sentence. A listener attempts to assimilate the meaning of each incoming word to his current semantic representation of the sentence, or indeed to his representation of the discourse as a whole. He may well rely upon his knowledge of the likely meaning of a sentence and a variety of heuristic parsing strategies. Perhaps these two activities are conducted in parallel (cf. 14). But, if they are both geared to a semantic goal, it seems more likely that they will be conducted in serial fashion in order to avoid conflicts between them. Studies carried out by Rosenberg and his colleagues (e.g. 151) have indeed shown that general semantic expectations affect the intelligibility of sentences presented in noise. Forster & Ryder (58) have observed comparable effects upon rapid serial visual perception. Just as sentences with a complicated syntax were harder to perceive accurately than sentences with a simple syntax, so also sentences describing improbable situations were harder to perceive accurately than sentences describing mundane situations. However, when complex syntax was combined with improbable content, the resulting sentence was still harder to perceive, compatible with the view that hypotheses about sense and syntax are pursued serially.

To deny the perceptual reality of deep structure is tantamount to denying the reality of surface structure, since some admixture of both sorts of information must be involved in syntactic processing. In what situation might it be shown that individuals are sensitive to unadulterated surface structure? One answer is, of
course, whenever there is enough time for its mental derivation. There are a number of studies that amply fulfil this requirement, yet which have failed to establish its psychological reality (cf. 111). Edwin Martin (118) found that when individuals were asked to sort the words of a sentence into "natural groups," their performance did not reflect the surface structure of the sentence. A comparable failure occurred in another study (120) in which subjects indicated the locations of intonation breaks in sentences. An interesting finding occurred in both of these studies: when the object of a sentence was a complicated nounphrase, it was often separated from the verb. Once again, it is possible that performance reflected semantic rather than syntactic processing.

THE COMPREHENSION OF WORDS

The comprehension of words is a prerequisite to the comprehension of sentences; and the meanings of words must be represented in memory in a way which is compatible with the meanings of sentences. Yet such is the profusion of hypotheses about meaning (e.g. 103, 157, 165), that thinkers of a positivistic persuasion might be tempted to argue that the notion is gratuitous. This position is advocated by certain contemporary philosophers (cf. 44), but it is unlikely to find much support among psycholinguists, though the drift of a recent work is at times in this direction (133). The more dominant tendency is to search for a parsimonious theory of the representation of meaning.

The Organization of Semantic Memory

The representation of a word in a mental dictionary must include a variety of information—its sound and orthography, its syntactic properties, its meanings and any relevant perceptual features, its emotional and factual connotations. The list is long, and granted the efficiency with which the information can be retrieved, the system must be highly organized (127). The heart of the problem for our purposes is the organization and retrieval of semantic information. But before the meaning of a word can be retrieved, it is first necessary to identify the word. It would take us too far afield to consider the vast literature on this process; and we note in passing only that the effect of even an apparently simple variable such as frequency of usage is a matter of considerable controversy (e.g. 29, 65, 108, 131).

Empirical studies of semantic memory are currently concerned with establishing the relative ascendancy of the psychological notion of an association, the linguistic notion of a semantic component, and the programming notion of a list structure. The traditional conception of an association has been reanimated by the work of Quillian (e.g. 147) in developing computer programs for the storage and retrieval of verbal information. Although this work was originally conceived as an exercise in artificial intelligence, Collins & Quillian (37, 40, 41) have used it as the basis for a model of human performance. They assume that the meaning of a word is its set of verbal associations, which may involve a variety of different sorts of associative link, including class-inclusion and part-whole relations. The organization of the complete associative network reflects the way in which information is initially
acquired. An individual is unlikely to learn as a separate fact that, for example, poodles are animals; rather he learns that dogs are animals and, independently, that poodles are dogs. Hence, the representation of "poodle" is only indirectly linked to the representation of "animal"; and once the network has been entered, the retrieval of certain facts is direct (e.g. "dogs are animals"), whereas the retrieval of other facts requires additional associative links to be traversed (e.g. "poodles are animals"). The results of several experiments suggest that the network can be entered with equal ease at any point (66, 113, 114), and they support the notion of a hierarchical retrieval of facts (37, 39, 40, 161). One unexpected finding was that although the closeness of items in the semantic network facilitated positive judgments about them, it impeded negative judgments, e.g. it is harder to evaluate "a canary is an ostrich" than to evaluate "a canary is a fish." The difference has been consistently replicated (155, 156, 182) and perhaps constitutes the central regularity of semantic judgment: the greater the similarity in meaning between words, the easier it is to make a positive judgment, and the harder it is to make a negative judgment, about a semantic relation between them. The network model provides a natural account for the facilitation of positive judgment, but it is necessary to suppose that negative judgments are impeded by extraneous links between neighboring words (40).

The regularity is, however, open to alternative explanations. Schaeffer & Wallace (156) argue that it is a consequence of a comparison process rather than a retrieval process. The meaning of a word, according to this theory, is represented by a set of conceptual elements analogous to the semantic components of linguistic theory (101). In deciding whether an item is a member of a category, the sets of conceptual elements for both words are retrieved and compared. The comparison involves sampling a certain amount of information from both sets in order to satisfy a criterion of judgment. Since similar words will have some common elements, the sampling required to satisfy the criterion will be smaller for a positive judgment than for a negative judgment. A still simpler explanation of the facilitation of positive judgments is that an individual has to search through a list of items stored with the category label, and so the larger the size of the category, the longer it will take him. Such category size effects were discovered by Landauer & Freedman (106), and subsequently corroborated by others (e.g. 122). They may explain what appears to be an effect of an associative hierarchy simply because position in this supposed hierarchy tends to be confounded with category size. It may take longer to decide that a poodle is an animal than to decide that it is a dog because there are more animals than dogs to be examined.

Perhaps such a result simply reflects Marbe's classical law of associations. The law states that the more frequent the associative response to a given word, the faster it should be made; and presumably "poodle" is more likely to elicit "dog" than "animal." The existence of associative norms for categories (e.g. 9, 115) has led to direct tests of the law, and it has been emphatically sustained in studies of both the production of category members (66, 113, 116) and the categorization of given items (182). But what happens when associative frequency is controlled? A study by E. E. Smith and his colleagues (161) paired arbitrary digits with a hierarchy of categories and still detected hierarchical effects. Comparable results were obtained in other
more naturalistic studies conducted by Carol Conrad. She found that when associative frequency was held constant, hierarchical effects were present for statements about class inclusion but absent for statements about the properties of items (42). There seems to be more than Marbe's law at work in categorization tasks; and of course it could never really explain the phenomena since it predicts that any item with a high associative frequency should be given as an instance of a category. The law is only a statement of a correlation between frequency and latency, and it cannot account for adherence to the logic of categorization. This failure brings us back to the three main competing approaches to semantic judgment. The attempts to tease them apart have been curiously unsatisfactory. No one theory has been conclusively vindicated, and they have all been confronted with some ugly facts.

Perhaps the major inconvenience for the hierarchical theory is a finding reported by Meyer & Ellis (cited in 107). They examined the times taken to categorize a variety of items including pronounceable strings of letters that were not actual words in the language. It took longer to decide that an item like “mafer” was not a “structure” than to decide that it was not a “building.” Because “mafer” is not a word, it cannot be part of a semantic network, and the hierarchical theory cannot explain the result. The category size hypothesis, however, runs into comparable troubles. Collins & Quillian (38) investigated negative judgments involving items totally dissimilar in meaning and categories of different sizes. They found that it was just as easy to decide that an item such as “magnesium” was not an “animal” as to decide that it was not a “dog.” There were equally disquieting results with some positive judgments, e.g. there was no difference between categorizing birds and categorizing animals. Finally, Landauer & Meyer (107) have cast doubt upon the main regularity of negative judgments, and consequently upon both the hierarchical theory and conceptual element theory. They obtained rankings of the closeness of meaning of items involved in a categorization task. Negative judgments were not appreciably affected by the degree of semantic similarity between item and category.

An empirically minded critic is likely to argue about the adequacy of these experiments and to try to draw up a balance sheet assessing their respective merits. Certainly there is a need for better data. No study appears to have utilized an entirely proper procedure for estimating category size, i.e. the number of items individual subjects actually recognize as members of a category. Yet a disinterested observer might draw a different moral. The seemingly contrary findings may reflect the flexibility of human semantic processing and the capability of subjects to adopt different strategies to meet the constraints of experimental situations. The experiments may even pass by the fundamental problem, since it is so rare in everyday life to have to ascertain the truth of a sentence such as “a dog is an animal.” Of course, one can hardly understand such a sentence without verifying it. But there is a considerable distinction between these two transactions for ordinary contingent sentences like “a dog is expensive to keep” (cf. 152). In understanding this sentence, an individual presumably retrieves some information from the lexical entry for “dog,” but what? It seems that for the most important semantic retrieval problem, we have no clear idea of what is retrieved, let alone how it is retrieved.
The Meaning of Words

It would be dangerous, even if it were possible, to decide between the competing theories of semantic memory solely upon the results obtained from studies of categorization. Other experimental tasks have, of course, been used to investigate lexical memory. They range from studies of judgments about whether or not strings of letters are words (e.g. 124, 153, 163) to procedures involving memory for words (e.g. 5, 74, 178). But these studies have been concerned mainly with matters other than a systematic assay of the alternative theories of semantic memory. From this point of view it is particularly unfortunate that the categorization studies have concentrated on the organization of nouns, since in them the distinction between linguistic and factual knowledge is especially blurred. On what basis, for instance, do we assign things to the category of "birds"? Are the criteria linguistic or factual? Indeed, perhaps there are no criteria, but merely a list of exemplars from whose shifting attributes a disjunctive set of family resemblances may be inferred. Yet some characteristics are more salient than others. Eleanor Heider (81) has shown that apparently there are different degrees of category membership. Her subjects considered that, for example, robins and eagles come closer to the ideal bird than do chickens and ducks. And, as the linguist George Lakoff has pointed out, we appear to recognize a variety of different sorts of criteria whenever we hedge our remarks in such characteristic ways as: "A robin is a typical bird"; "Strictly speaking, an ostrich is a bird"; "Superman is a regular bird."

Many of these complexities melt away when one turns to verbs or to relational terms in general. Apart from technical vocabulary, their meanings rarely embrace a fuzzy boundary between linguistic and encyclopedic knowledge, but the relations between them are correspondingly more complex. Simple hierarchies give way to intricate cross-classifications; for example, such verbs as "travel," "own," and "see" lack a causal component that is present in the meaning of "lift," "seize," and "look at," and these verbs in turn lack an intentional component that is present in "chase," "buy," and "look for." An alternative and more salient classification of these examples, which are taken from an unpublished analysis by George Miller and the present author, is simply into verbs of motion, verbs of possession, and verbs of perception. It is unlikely that mere lists of verbs in semantic memory could handle these subtleties. And since the network theory has no way of freeing meaning from the specific verbiage in which it is expressed, it seems committed to the dubious view that the meanings of verbs are acquired, in part, from explicit statements involving "cause" and "intention." Only a theory countenancing underlying conceptual elements seems to offer sufficient flexibility to account for the acquisition of meaning. Apart from such global considerations, however, it is difficult to distinguish between a semantic network and a set of conceptual elements. Studies of semantic judgment often yield evidence compatible with both theories. For instance, Steinberg (e.g. 164) found that sentences like "The husband is a wife" are considered to be contradictory, whereas sentences like "The chair is a husband" are considered to be nonsensical. The difference can be explained either by postulating an ordered set of semantic components representing the meaning of a word or by reference to the likely distance between words in the semantic network.
One factor that has undoubtedly shaped the direction of research into semantic memory, at least in a negative way, has been a lack of systematic knowledge about the meanings of words. It is difficult to study the organization and retrieval of information without a specification of what information is involved. Although a number of empirical techniques are helpful (e.g. 54, 136), there seems to be no substitute for painstaking linguistic analysis. Unfortunately, linguists have been reluctant to carry out such lexicographical projects in the detail required by psychologists. Perhaps the sort of approach that will have to be adopted is exemplified by George Miller's recent study of verbs of motion (128). Miller used a novel method of "incomplete definition" in which his intuitive choice of paraphrases revealed the major semantic elements of each verb. The resulting analysis was tested empirically at various points by using a technique in which a subject has to sort words according to his perception of their similarity in meaning. This technique has been developed over a number of years for use with both adults and children (e.g. 3, 126), since it provides a stringent test of semantic analyses. Once a beachhead into the English lexicon has been established in this way, it may be possible to devise suitable discovery procedures.

THE COMPREHENSION OF SENTENCES

A major obstacle to understanding the process of comprehension is its intangible nature—a fact borne out by the possible semantic contamination of studies of syntactic perception. It can, of course, be investigated by the use of such tasks as verification or paraphrase, since—to use a potent metaphor—a program must be compiled before it can be executed (45). But the way in which comprehension occurs may well depend upon which of these tasks is about to be performed. Perhaps the simplest and oldest technique is merely to time how long it takes an individual to decide that he has understood a sentence. It was with this procedure, for example, that Danks (43) showed that comprehension appears to be disrupted more by semantic anomaly than by grammatical deviance, and that the effort after meaning sometimes led subjects to misread lexical items. It is probably important whether the subjects read or listen to the material. A reader, unlike a listener, can adjust the depth of his processing to take into account fluctuations in the difficulty of a passage (46). Thus readers can make rapid and accurate assessments of the comprehensibility of a sentence, sometimes even before they have reached the end of it (158). A listener can, of course, elect not to pay too much attention to what someone is saying, but there is clearly no aural equivalent to rapid reading—it is speech that has to be speeded up to increase the efficiency of the system (63).

There is a direct parallel between judgments of comprehensibility and judgments of grammaticality; both sorts of judgment appear to be adversely influenced by semantic incongruities between noun and verb (e.g. 130, 170). The crucial syntactic factor appears to be simply the distance between subject and verb; hence sentences that are self-embedded, or that have a considerable Yngve depth, tend to be rated as hard to understand (79, 171). It is difficult to establish whether the nouns are
more salient than the verb, or the verb is more salient than the nouns. But the verb does appear to be decisive for the control of the interpretative process; and when sentences are grouped together on the basis of their similarities in meaning, sameness of verb rather than sameness of subject is the potent factor (cf. 80).

Understanding Ambiguous Sentences

Most sentences considered in isolation are to some extent ambiguous; but it is only rarely that listeners are perplexed by ambiguity. In some mysterious way, the context of an utterance usually suffices to render its meaning plain. However, although only a single conscious interpretation of a sentence occurs, other interpretations may have been covertly considered and rejected. Certain results support such a conjecture; for example, Foss (60) found that responses in his phoneme-monitoring task were slower to ambiguous than to unambiguous sentences, and Garrett (67) found that the location of clicks within ambiguous words was less accurate than when they had been rendered unambiguous by the preceding context. But perhaps the most striking evidence for the partial processing of more than one meaning comes from a study by Lackner & Garrett (105). They presented an ambiguous sentence to one ear with a disambiguating sentence 5 dB softer to the other ear. Almost without exception the sentence was interpreted in accordance with its disambiguating context, yet the subjects were unable to report this unattended sentence. In order for the context to have such an effect, it seems that both meanings of the sentences must be to some extent available to the subjects. All of these findings lend support to McKay’s theory (117) that the meanings of ambiguous sentences are processed in parallel and interact with one another in such a way that the perception of one meaning leads to the suppression of others.

Failures to detect effects of ambiguity have been reported in a number of studies. It seems to take no longer to paraphrase ambiguous sentences than unambiguous sentences (67); and it only takes longer to verify ambiguous descriptions on those occasions that the ambiguity is likely to have been noticed (27).

These superficially inconsistent findings have been resolved by Garrett (67) on the grounds that the experiments showing no effects of ambiguity tested subjects after they had understood the sentences, whereas experiments showing effects of ambiguity tested them during the process of comprehension. The one awkward phenomenon is the greater time required to complete an ambiguous rather than an unambiguous sentence fragment (117). However, a recent study (17) suggests that the effect appears only when the fragment ends in an incomplete clause. In other words, ambiguity is irrelevant once a clause boundary is passed, because a single meaning has generally been decided upon by then. There is, however, an alternative resolution of the experimental findings. An individual is likely to analyze more than one meaning of a sentence only if its context fails to provide a strong bias towards one reading. If there is such a bias, the sentence will be no harder to understand than an unambiguous one. If the bias turns out to be inappropriate, then the listener will have been led up the garden path and he will have to reconsider the sentence (cf. 47). It follows that where sentences are presented in contextual isolation, effects
of ambiguity are likely to be detected; but where they are presented after a disambiguating context, no such effects are likely to be detected. The one awkward exception to this principle is the failure to find any difference in the time to paraphrase isolated ambiguous and unambiguous sentences. But although paraphrasing tasks are often revealing about the search for meaning (51, 68), they are relatively insensitive indices of perceptual processing (76). It seems, in general, that any procedure permitting sentences to be considered in tranquillity is unlikely to be affected by their ambiguity (129).

Students of ambiguity have been provided with a wide variety of materials by linguists, and they have investigated the effects of ambiguities in both words and syntax. It does not seem that there are any safe generalizations that can be made at present about the psychological relevance of the origin of an ambiguity in a sentence. However, the view that all the meanings of a sentence are analyzed to some extent seems more feasible when its ambiguity is structural rather than lexical in origin; certainly such a distinction would explain why it is harder to reinterpret sentences with lexical ambiguities (cf. the study by Cairns cited in 17).

The Verification of Sentences

The study of sentence verification has undergone a distinct shift of emphasis during the last few years. Originally the task was introduced in order to establish differences in the ease of understanding sentences, as in Wason's classic studies of negation (e.g. 172). But psycholinguists have become increasingly interested in how an individual determines whether a sentence is true or false. This development is due largely to the independent work of Clark (32, 36) and Trabasso (167, 168), although the differences between their approaches are largely terminological. Both theories agree that the process of verification can be separated into four main serial stages: the representation of a sentence, the representation of some state of affairs it purports to describe, the comparison of the two representations, and the mobilization of an appropriate response. The order of the first two stages depends upon the order of presentation of the materials or the order in which they are encoded. But this order has important consequences. If a visual display is presented first, it will usually be represented in a neutral affirmative way. It it is presented after its purported description, then the format of its representation will be contingent upon the nature of this sentence.

Verification tasks have shown that a variety of factors appear to influence the initial representation of a sentence. There is evidence that negative sentences take longer to understand than affirmative sentences (32, 168), that explicit negatives such as "none" or "not present" take longer to understand than implicit negatives such as "few" or "absent" (32, 99), and that passive sentences take longer to understand than active sentences (135). But the analysis of verification has also stimulated interest in other linguistic phenomena, including the distinction between unmarked items, such as "deep" or "long," which can be used in both a neutral and a contrastive sense, and marked items, such as "shallow" or "short," which can be used only in a contrastive sense. Marked items resemble negatives in that they tend to take longer to encode. However, there has been sharp controversy between
Huttenlocher (e.g. 87, 88) and Clark (e.g. 33) about lexical marking and the strategies individuals use to solve three-term series problems. The controversy has been fervent rather than enlightening, perhaps because the antagonistic positions may be reconcilable (96).

Although Clark (e.g. 35) originally argued that sentences are represented in the form of their underlying structures, the evidence for this view is sketchy and indirect and takes no account of the alternative hypothesis that they are represented in a more semantic form. No verification study has directly contrasted these two views; but evidently Clark (34) has partially abandoned this tenet of his theory.

There is also a latent controversy about possible differences in picture encoding. Chase & Clark (31) argue, for example, that whether a display is encoded as \( X \) above \( Y \) or as \( Y \) below \( X \) makes no difference to the duration of the operation. They found that instructing subjects to attend to the top or the bottom of the display had no effect on encoding time, though it obviously affected the form in which the display was represented. However, when Olson & Filby (135) ingeniously focused attention upon the depicted object of an action, encoding was in a form comparable to the passive voice and took longer to execute. It seems that verification like other choice reactions is influenced by information-theoretic aspects of the task, and the range of alternative stimuli may affect encoding time (134).

The core of the verification models is, of course, the stage when the two representations are compared. It is axiomatic in both models that a successful match depends upon complete congruency between representations; and this view naturally suggests an underlying abstract encoding common to both pictures and sentences. The challenge to the theorist is to devise a plausible set of operations that yields the appropriate responses and tallies with the data. Clark & Chase (36) argue that the comparison stage has a truth index which is initially set as "true," and that each mismatch reverses this index. A series of comparisons is required because the negative component of a sentence is not considered until after its basic unnegated proposition. The truth index obviously remains unchanged for a true affirmative. It is reversed once for a false affirmative because the representation of the sentence fails to match the representation of the picture. It is reversed once for a false negative because although its unnegated proposition matches the representation of the picture, its negative component does not. And it is reversed twice in the case of a true negative because first the unnegated proposition and then the negative component fail to match the representation of the picture. The resulting interaction between the form of the sentence and its truth value has been abundantly confirmed: true affirmatives are easier than false affirmatives, but true negatives are harder than false negatives (e.g. 32, 168). In some experiments, however, subjects rid themselves of negative sentences by converting them into equivalent affirmatives, e.g. the predicate "isn't odd" is changed to "is even." Naturally, such conversions are particularly prevalent with binary predicates (168). When subjects adopt such a strategy, their results are compatible with the models on the assumption that conversion modifies the representation of a sentence and takes an additional amount of time.

Contrary to the central assumption of the verification models, Tversky (169) has suggested that a name may be converted into a pictorial code or a picture into a
verbal code; and Seymour (159) has similarly argued that the meaning of a word in a shape-naming task may be represented in some quasipictorial fashion. Such arguments may be correct for simple tasks where subjects are sensitive to the restricted nature of the linguistic materials. But if negative sentences were encoded pictorially, it would be extremely difficult to explain the standard interaction with truth value. Perhaps the real burden of these criticisms is that the boundary conditions of the verification models have yet to be clearly drawn. The majority of studies have been concerned with seeking confirmatory evidence, and critical tests of the serial nature of verification, using the additive factor method (166), have only recently been undertaken. Unfortunately, it does not seem too difficult to discover embarrassing interactions between factors, e.g. encoding a negative sentence appears to retard both picture encoding (104) and the retrieval of information from long-term memory (123).

**Answering Questions**

Answering a question is in many ways a comparable task to verifying a statement, especially when the interrogated information has only just been presented. Hence it is a simple matter in principle to modify the verification models so that they will account for this sort of question answering (35). But the actual details of such modifications are harder to establish since the experimental results are complicated. A surprising phenomenon was discovered by Patricia Wright (185) in a task where listeners heard a sentence followed 5 seconds later by a question about it. The main source of difficulty in giving a correct answer about the actor or object was a mismatch between the voice of the question and the voice of the original sentence. Consequently, passive sentences created difficulty only when the questions were active. When the action itself was queried—e.g. “what was done by X?”—the syntactic form of the question was much less relevant than which noun occurred in it. The task was easier when this noun was the first to be mentioned in the original sentence. It might seem that these effects are simply due to the verbatim retention of the sentences. In a subsequent replication, however, Wright (186) used an interpolated task to prevent rehearsal, but the pattern of results was identical. A further experiment suggested that even when the answer to a question is a single word, its production covertly involves a complete clause, and that the form of the question determines the form of this clause. This principle applies even where the action itself is queried: the noun that occurs in the question becomes the first noun of the answer.

There are still more complex results from studies of question answering. Some experiments conducted by Smith & McMahon (162) detected direct effects of voice: where the sentences and questions concerned the relative order of two participants, passive sentences took longer to encode than active sentences. There were also striking effects of both intricate semantic variables and the order of presentation of sentence and question. What makes such results particularly difficult to interpret is, as the authors acknowledge, the possibility that subjects adopted special strategies to deal with the task.
The Natural History of Negation

The main danger with the models of verification is that their very success in certain situations may lead to unjustified claims, such as that the comprehension of all forms of negation can be accounted for by a single model. The example of negation is instructive because Wason has argued for a number of years that its proper function is to deny misconceptions and misunderstandings. The reason that negative sentences are not ordinarily noted for their difficulty is perhaps because their interpretation requires the listener merely to consider changing the truth value of some proposition that he has already interpreted (174). Indeed, the ordinary negative seems analogous to the false negative of the laboratory, since it probably calls for only a single mental reversal of truth value (173). In evaluating a similar thesis, Greene (72) found that it was much easier to determine that an affirmative and a negative sentence differed in meaning than to determine that they were synonymous. It is also easier to appreciate that an assertion has been denied when the denial is an explicit negation rather than an implicit negation (98).

One tentative moral that might be drawn from this aspect of negation is that there are good psychological reasons for the existence of each sort of grammatical clause. If there are circumstances where negatives are easier than affirmatives, then there are likely to be favorable circumstances for any option within a clause. There may be no intrinsically difficult clause constructions, only difficult combinations of them.

LANGUAGE AND MEMORY

It is parsimonious to assume that sentences are remembered, at least initially, in the form in which they are interpreted; and such an assumption is often made, as Fillenbaum (53) has emphasized, when making inferences about comprehension from the results of memory experiments. But since these results so often depend upon the task the subjects have performed (26, 50, 187), there is uncertainty about whether the experimental procedure affects initial interpretation or subsequent memorial processes. Nevertheless, it is generally accepted that sentences are not normally retained as strings of unrelated words, and a variety of hypotheses have been put forward about their mode of organization in memory.

Memory and Syntax

A simple proposal is that sentences tend to be retained in "chunks" that correspond to their surface constituents (e.g. 94). Although effects of surface structure have been reported in the rote learning of connected discourse (4), their natural locus appears to be the short-term retention of sentences (82). When an individual is presented with a sentence spoken in normal intonation followed by a probe word to which he has to respond with the next word in the sentence, the latency of his response appears to reflect the surface structure of the sentence (1). Since pauses in linguistic and nonlinguistic materials have a comparable effect on the probe task (180, 181), the apparent effects of surface structure may be due to rhythmic aspects of the perceived or rehearsed intonation of the sentence (cf. 120, 121).
We saw previously that certain aspects of syntax affect comprehension, and several theorists have proposed that such features will also affect retention. A diversity of syntactic indices have been canvassed, including transformational complexity, Yngve depth, and lexical density (i.e. the proportion of content words in a sentence). Unfortunately, while Perfetti (142) found an effect of lexical density but no effect of depth in short-term retention, Bacharach & Kellas (7) holding lexical density and depth constant found an effect of transformational complexity, and, in more long-term tasks, Wearing (e.g. 176) found an effect of depth but little or no effect of transformational complexity. There is not likely to be any simple resolution of these results. It is difficult to see why any syntactic property should ordinarily be relevant to the retention of intelligible sentences, unless it is being deliberately used as an aid to memory. Paradoxically, unraveling the mechanisms of intentional memory could prove extremely difficult simply because of an enhanced sensitivity of subjects to the nature of the materials.

The obvious alternative to some sort of superficial segmentation is, of course, a deep structure representation of sentences. But since our earlier analysis of perception suggested that deep structure had no independent psychological validity, it should not be surprising to learn that recent work continues to cast doubt on this hypothesis. It was originally argued, for example, that the amount of storage space occupied by a sentence depended on the number of transformational “footnotes” such as Passive, Interrogative, and Emphatic specified in its underlying structure. This view has encountered a whole series of setbacks (e.g. 49, 61, 69). Deep structure theories about the efficacy of words as prompts to the recall of sentences have perhaps fared a little better, but they have not gone unchallenged. Certainly prompted recall can be affected by factors other than syntax. It is enhanced, for instance, by the presence of co-referential items in separate clauses (110), provided that the relation is conspicuously signposted by the use of a pronoun. The facilitation is not simply due to the greater ease of processing pronouns (119). On the contrary, children’s memory is improved by explicit repetitions of noun phrases within the same sentence (100), presumably because they have yet to master the intricacies of pronominalization. The integrative function of pronouns is further corroborated by the observation that when a sentence is partially recalled, the presence of a pronoun rather than a noun is more likely to lead to its complete recall (10).

Perhaps the best illustration of the drift away from explanations in terms of deep structure is provided by the work of Rohrman and his colleagues on memory for nominalizations. Linguistic analysis shows that a subject nominalization such as “growling lions” has a simpler underlying structure than an object nominalization such as “digging holes,” because the object nominalization requires an additional argument representing its deleted subject. Rohrman’s original view (cf. 149) was that this difference explained why subject nominalizations were easier to remember than object nominalizations. However, it was evident in the experimental materials that the subject nominalizations generally involved animate nouns and intransitive verbs, whereas the object nominalizations generally involved inanimate nouns and transitive verbs. Subsequent tests showed that animate nouns were easier to remember than inanimate nouns, and that intransitive verbs were easier to remember than
transitive verbs (144, 149). Rohrman concluded, with some justice (8, 150), that the critical factor was the lexical representation of nouns and verbs.

Just when it looked as though a plausible account of linguistic memory might be given in terms of undifferentiated syntactic factors on the one hand and semantic factors on the other hand, a quite separate conception of memory emerged.

**Memory and Imagery**

In a number of publications (e.g. 137, 138), Paivio has argued that imagery is a highly potent mode of representing sentences and other verbal materials. According to this hypothesis, certain so-called "concrete" sentences are represented in the form of sensory images, whereas other so-called "abstract" sentences are represented less efficiently in some form of verbal code. Particular situations are likely to inhibit the use of imagery (6), and particular individuals are likely to be incapable of forming sensory images (141). Nevertheless, imageability seems to be an important variable. Two independent studies, for example, have shown that Rohrman's subject nominalizations were rated as easier to image than his object nominalizations (139, 175); and when imageability was carefully manipulated, it was the only variable to have a consistent effect upon recall. Yet the interaction between imageability and type of nominalization, with its varying pattern from one experiment to the next, suggests that unknown factors are also at work.

There have been several other reappraisals of established findings in the light of the imagery theory. It was generally considered that the subjects of sentences are easier to recall than their objects, and that they differ in their efficacy as prompts in the same way (86). However, when imageability was controlled, these differences disappeared (89). Likewise, earlier studies had shown that the nouns in adjective-noun pairs tend to be better recalled than the adjectives (86), and that they differ in their efficacy as prompts in the same way (112). But when their imageability was equalized, the asymmetry in recall disappeared (12, 143).

It seems incontrovertible that sentences can be remembered in the form of images. But although such a view is often pitted against alternative linguistic explanations of experimental findings, it is not necessarily in fundamental conflict with them. When imagery is used, sentences have to be translated into images, and at recall the images have to be translated back again into sentences. The probability of error during these processes, as Paivio (139, 140) admits, may well be related to the complexity of the material to be remembered. Certainly linguistic factors seem to be necessary to account for the finding that the imagery value of the object noun is a better predictor of verb recall than the imagery value of the subject noun (89). And if the differences in rated imageability could be explained, they might depend upon linguistic complexity as well as more obvious perceptual attributes (102).

The task for the Imagist psychologists is to demonstrate conclusively that concrete sentences are encoded as sensory images, and that it is in virtue of this representation that they are better remembered than abstract sentences. Begg & Paivio (13) found that subjects were sensitive to changes in the meaning of concrete sentences, and merely sensitive to changes in the wording of abstract sentences. Such differences are consistent with the theory but inconclusive. A recent study has
demonstrated that the abstract sentences were harder to understand and that changes in their meaning were correspondingly harder to detect (92). Hence, the most that can be safely concluded from the original experiment is that individuals retain the sense of sentences, as opposed to verbiage, in proportion to their grasp of it. More direct tests of encoding have also sometimes yielded results contrary to the imagery theory (e.g. 179). It seems incontrovertible that sentences can be remembered without recourse to imagery.

One of the consequences of the revival of interest in imagery, and in other forms of mnemonic, has been an enormous increase in the use of sensible linguistic materials in studies of associative learning; and there is wide-ranging evidence for an improvement in such tasks as paired-associate learning when subjects are encouraged to form a unifying image between stimulus and response (e.g. 19, 20). The study of sentences from an associative point of view—a topic too vast to be reviewed here—has revealed that in incidental learning tasks a sentence has a definite integrative function (2). Likewise, when subjects respond associatively to one word with another from the same sentence, the pattern of their responses is consistent with neither a surface nor a deep structure representation. It is best explained in terms of semantic processes (177).

**Memory and Meaning**

When subjects are unaware that their memory for a particular sentence is going to be tested, they rapidly lose the ability to recall its precise wording. In a series of studies, Jarvella (e.g. 90, 91) has investigated the ability of listeners to perform a running memory task in which from time to time they have to recall the immediately preceding part of a connected passage. Typically, memory is reasonably accurate only for the final sentence that was heard, and it is verbatim only for its final clause. Caplan (25) has shown that the latency to identify a probe word is faster if it, too, comes from the immediate clause rather than an earlier clause, regardless of the amount of interpolated material. Similarly, a study carried out by Goldman-Eisler (70) has established the importance of the clause in simultaneous translation. Such results bear out the thesis that verbatim information is normally lost to memory almost as soon as it has been used, and that the clause is the natural unit in such transactions.

One of the difficulties in evaluating the importance of the clause in perception is to determine whether its boundary is associated with the assignment of underlying structure or with a more general process of semantic interpretation. Its importance in memory is similarly ambiguous since the retention of meaning is often sufficient to enable a subject to reconstruct underlying structure. However, where sense and syntax are not confounded, it appears that underlying grammatical relations are rapidly forgotten. Fillenbaum (53) found that subjects in an incidental memory task readily confused sentences of the form “If you do that I'll hit you” with those of the form “Do that and I'll hit you.” Another recent study obtained similar confusions between sentences of the form “John liked the painting and he bought it from the Duchess” and “The painting pleased John and the Duchess sold it to him” (97). The meaning of sentences in these studies and others (e.g. 11) was retained with
remarkable accuracy. And, as Rosenberg and his colleagues have shown, purely semantic manipulations can affect the memorability of sentences. Semantically well integrated sentences such as "The doctor cured the patient" are better recalled than less well integrated sentences such as "The doctor shook the author" in both intentional and incidental tasks (e.g. 151).

Inference is likely to play an important part in comprehension, and it is from studies of memory that this fact first emerged in the psycholinguistic literature. It had been forced upon theorists by some of the problems of analyzing connected discourse. For example, Winograd (184) pointed out that there is no linguistic rule for relating anaphoric pronouns to their appropriate noun phrases. The listener relies upon a heuristic inference in order to appreciate the contrast between "Peter put the package on the table, but because it was round, it rolled off" and "Peter put the package on the table, but because it wasn't level, it slid off" (cf. 132). The role of inference in memory has been most notably demonstrated by Bransford & Franks and their colleagues (21). They observed that subjects presented with a sentence like "Three turtles rested on a floating log, and a fish swam beneath them" would readily assume in a recognition test that the sentence had read "Three turtles rested on a floating log, and a fish swam beneath it." However, if the original sentence merely stated that the turtles were beside the log, such an inference is less plausible, and the subjects were much less likely to make the false recognition. Such inferences probably occur during the interpretation of a sentence; yet an unpublished study by Charles Jenkins (cited in 71) suggests, surprisingly, that they are made later.

Memory and Connected Discourse

It is debatable whether in the ordinary course of events individual sentences are remembered. It seems more likely that once their meaning has been grasped, they will be retained not as separate entities but by integration into the relevant stock of long-term memories. Bransford & Franks have shown that there appears to be a strong tendency for separately acquired information to become related in this fashion. Their subjects had considerable difficulty in distinguishing whether they had been originally presented with such sentences as "The ants ate the jelly" and "The jelly was on the table" or with more complex sentences such as "The ants ate the jelly on the table" (22, 64). A further study has established that such integrations occur in the spontaneous interpretation of connected discourse (93).

Psychologists have been interested for a long time in the ways in which coherent discourse is remembered, since it is a problem of both practical and theoretical significance (cf. 28 for a number of reviews from both orientations). The main topic that has been investigated experimentally is the effect of thematic knowledge upon retention. One of the first of these studies, carried out by Pompi & Lachman (145), showed that where the theme of a text was obvious, subjects would readily assume that words related to it had occurred in the passage. Dooling & Lachman (48) went on to demonstrate that the recall of a somewhat vague passage was considerably enhanced when it was given an appropriate and revealing title. A similar study by Bransford & Johnson (23) replicated such findings and also showed that the comprehension and retention of sentences could be improved by presenting a picture of the
general situation they described. Few psychologists would be surprised by this result because without the picture to set the scene the sentences hardly appear to relate to one another—the reader is distinctly reminded of the White Rabbit's evidence in *Alice in Wonderland*. What is particularly interesting about such passages, and stands in need of further investigation, is the variety of linguistic devices that can be exploited to avoid having to make clear what is being discussed, e.g. the use of anaphoric pronouns in the absence of a prior specification of their reference, e.g. “It takes some skill but it's easy to learn,” the use of abstract or generic descriptions for specific activities, e.g. “Too many people doing the same thing can also cause problems,” and the deliberate omission of an underlying argument, e.g. “Rain, however, soaks in very fast.” (All of these examples come from a passage about flying a kite.) It seems plausible that such variables would have a similar effect upon the memory of individual sentences; and, indeed, Dooling & Lachman (48) found such effects even when the passage was presented in a scrambled order.

The ultimate problem in the study of connected discourse is to discern what factors make for its maximum cohesion and how it is mapped into some form of memorial representation. We know that individuals tend to remember what is important (e.g. 95), and that they are sensitive to structural and cohesive factors in the text (e.g. 146), but we are almost totally ignorant about the processes involved in the long-term representation of discourse. We do not even know whether such linguistic memories entail different principles of organization from the episodic memories of daily life.

**Literature Cited**


60. Ibid 1970. Some effects of ambiguity upon sentence comprehension. 9:699–706
EXPERIMENTAL PSYCHOLINGUISTICS

105. Lackner, J., Garrett, M. F. Resolving ambiguity: effects of biasing context in the unattended ear. *Cognition*. In press
111. Levelt, W. J. M. 1970. A scaling approach to the study of syntactic relations. See Ref. 15. 109–21


134. Olson, D. R. 1972. Language use for communicating, instructing, and thinking. See Ref. 28, 139–67


145. Pompi, K. F., Lachman, R. 1967. Surrogates processes in the short-term re-
perceptual reality of the phoneme. 
152. Rubenstein, H. An overview of psycholinguistics. See Ref. 35
156. Ibid 1970. The comparison of word meanings. 86:144–52


